PRITHVIRAJ CHAVAN

+91 7795844029 | Bangalore, India
chavanprithvi45@gmail.com | linkedin.com/in/prithvi-raj-chavan | github.com/Prithvi-ra-j

SUMMARY

Data Engineer with expertise in building scalable data pipelines, ETL workflows, and distributed data processing using Python, SQL, and Apache Spark. Proficient in relational databases (PostgreSQL, MySQL) and machine learning model deployment. Adept at collaborating with cross-functional teams to deliver data-driven solutions, with a focus on data integrity and performance optimization.

EDUCATION

Bachelor of Engineering, Computer Science and Engineering 2020 – 2024
K.S. School of Engineering and Management, Bangalore

SKILLS

Programming	Python, SQL, Java
Data Engineering	Apache Spark, ETL, Data Pipelines, Feature Engineering
Distributed Systems	Parallel Processing, Data Partitioning, Scalability
Machine Learning	Scikit-learn, PyTorch, NLP, Model Deployment
Databases	PostgreSQL, MySQL
Tools	Pandas, Power BI, SQLAlchemy, Git
Concepts	Data Integrity, Data Warehousing, EDA
Soft Skills	Collaboration, Problem Solving, Analytical Thinking

EXPERIENCE

Machine Learning Intern Aug 2023 – Sep 2023
Prinston Smart Engineers, Bangalore

- Built a telecom churn prediction model with 90% accuracy using customer data, supporting scalable retention strategies.
- Automated ETL pipelines and feature engineering workflows, improving data processing efficiency by 20%.
- Conducted EDA and model validation, reducing training time by 15% through optimized workflows.
- Collaborated with stakeholders to deliver Power BI dashboards, enabling data-driven decisions.

PROJECTS

E-Commerce Data Pipeline (UCI Online Retail Dataset)

- Designed a scalable data pipeline to process e-commerce customer data (500k+ records) using Python and Apache Spark for distributed processing.
- Developed a PostgreSQL schema and implemented ETL workflows with SQLAlchemy, handling 135k missing CustomerID values to ensure data integrity.
- Engineered metrics like purchase frequency and trends to support business analytics and decision-making.

Sepsis Prediction Model (MIMIC-III Dataset)

- Built a machine learning model to predict sepsis onset in ICU patients, handling large-scale time-series data (40k+ records) for improved scalability.
- Engineered features from clinical vitals, achieving an 8% reduction in estimated mortality risk through early detection.

CERTIFICATIONS

Machine Learning with Apache Spark – IBM Coursera (2025)
IBM Data Science Professional Certificate – Coursera (2024)
Python Certification – HackerRank (2024)
AWS Cloud Technical Essentials – Coursera (2025)